

First-principles photoemission spectroscopy of DNA and RNA nucleobases from Koopmans-compliant functionals

Ngoc Linh Nguyen,^{1,*} Giovanni Borghi,^{1,2} Andrea Ferretti,² and Nicola Marzari¹

¹*Theory and Simulations of Materials (THEOS),
and National Centre for Computational Design and Discovery of Novel Materials (MARVEL),
École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*
²*Centro S3, CNR-Istituto Nanoscienze, 41125 Modena, Italy*

The need to interpret ultraviolet photoemission data strongly motivates the refinement of first-principles techniques able to accurately predict spectral properties. In this work we employ Koopmans-compliant functionals, constructed to enforce piecewise linearity in approximate density functionals, to calculate the structural and electronic properties of DNA and RNA nucleobases. Our results show that not only ionization potentials and electron affinities are accurately predicted with mean absolute errors < 0.1 eV, but also that calculated photoemission spectra are in excellent agreement with experimental ultraviolet photoemission spectra. In particular, the role and contribution of different tautomers to the photoemission spectra are highlighted and discussed in detail. The structural properties of nucleobases are also investigated, showing an improved description with respect to local and semilocal density-functional theory. Methodologically, our results further consolidate the role of Koopmans-compliant functionals in providing, through orbital-density-dependent potentials, accurate electronic and spectral properties.

PACS numbers: 71.15.Mb, 74.25.Jb, 79.60.-i

I. INTRODUCTION

The nucleobases adenine (A), cytosine (C), thymine (T), guanine (G), and uracil (U) are the primary building blocks of deoxyribonucleic (DNA) and ribonucleic (RNA) acids. The sequence of base pairs, which are stacked upon one another leading directly to the helical structure of DNA and RNA, carries all the genetic information of living organisms. Due to such biological importance, a proper understanding of the structural and photoelectron properties of these molecules is a priority, in order to unveil the mechanisms of formation of DNA and RNA chain, and the reaction dynamics under exposure to ultraviolet light or ionizing radiation¹.

Extensive studies to understand the electronic properties of nucleobases have been carried out for many years, both theoretically and experimentally^{2–13}. Notwithstanding such efforts, many questions related to photoelectron properties, such as the stability and symmetry of ionized states, are yet to be understood^{5,14}. Furthermore, under experimental conditions, nucleobases appear in several tautomeric or conformeric variants, which differ from one another only in the position of a hydrogen in the structure, and which have energies lying very close to each other. This makes the detailed understanding of spectral properties quite challenging, since the contributions of each single isomer are hard to resolve. Distinguishability can be attained through photoemission experiments with high photon energies, such as X-ray techniques^{15,16}; these, however, are not suitable for examining the properties in vivo, and do not address the energies of the valence electrons. For these reasons, a synergy of low-energy experiments and accurate theoretical simulations would be most beneficial to interpret photoemission measurements.

Several *ab-initio* ground-state calculations have already been carried out in order to determine which tautomers are the most energetically favorable^{15,17}. Similarly, many efforts have also been devoted to predicting photoemission spectra^{5,7}. These efforts aim at understanding the nature of the spectral peaks, possibly labeling them with their respective symmetry quantum numbers.

From the theoretical point of view, most photoemission studies have been performed using many-body perturbation theory or high-level wave-function methods (see Ref. 11 and references within), whose considerable computational cost prevents them from being applied to more complex biological environments or to sets of paired bases. This is the reason why simpler methods such as Hartree-Fock (HF) or density-functional theory (DFT), computationally less demanding, are still frequently employed. Unfortunately, their accuracy in the calculation of ground state energies is not complemented by a comparable precision in predicting electronic excitation energies and photoemission spectra¹⁸. These drawbacks are intrinsic to Kohn-Sham (KS) DFT whose single particle energies [except for the highest occupied molecular orbital (HOMO)^{19,20}] cannot even in principle be interpreted as quasiparticle excitation energies^{18,21,22} (though arguments exist suggesting that exact KS eigenvalues may provide good approximations to them^{23–25}).

Recently, Koopmans-compliant (KC) functionals were introduced^{26–32} to enforce a generalized criterion of piecewise linearity (PWL) in the energy of approximate DFT functionals with respect to the fractional removal or addition of an electron from any orbital of the system. This PWL condition is a generalization to the entire manifold of the molecular DFT+U approach^{33,34}, stemming from similar linearization ideas in the solid state.³⁵ The

condition of Koopmans’ compliance is naturally akin to that of enforcing a correct description of charged excitations^{28,36}, and can therefore lead to orbital energies that are comparable to the quasiparticle excitation energies of photoemission experiments. In a previous work³⁶, we showed the remarkable performance of KC functionals in predicting ultraviolet photoemission spectra (UPS) and orbital tomography momentum maps for photovoltaic molecules, showing an agreement with experiments for frontier orbital energies [ionization potentials (IPs) and electron affinities (EAs)] that is comparable (in some cases even slightly superior) to state-of-the-art methods in many-body perturbation theory, while preserving a moderate computational cost and scaling³², and the quality of potential energy surfaces of the underlying DFT functionals³².

In this work we perform a study on DNA and RNA nucleobases using the best performing KC functional, labeled KIPZ³², whose performance on small molecules was assessed in Refs. 32 and 36. Accuracy in predicting spectroscopic properties of DNA and RNA bases is compared here to experiments, standard DFT calculations, many-body perturbation theory, and quantum-chemistry methods. We illustrate the effectiveness of the KC approach in distinguishing tautomers and in correctly predicting the geometrical properties of nucleobases (so far accessible only via second-order Møller-Plesset (MP2) perturbation theory³⁷).

The paper is organized as follows: in Sec. II we review the formulation of Koopmans-compliant functionals and provide a simple algorithm to perform self-consistent structural optimizations. In the second part (Sec. III) we study the binding energies of the frontier orbitals (Sec. III B) and the photoelectron properties and photoemission spectra (Sec. III C) of the most stable nucleobase tautomers. Finally (Sec. III D), we discuss molecular geometries and show the effects of structural changes on photoelectron properties.

II. METHODOLOGY

The main advantage of KS-DFT over many-body perturbation theory techniques or quantum chemistry methods is that the electronic ground state density of a system is inexpensively parametrized through a set of single-particle orbitals which are the result of the diagonalization of an effective noninteracting KS Hamiltonian. This enables to carry-out calculations that, whilst including many-body effects through the density functional, have the same computational cost of an independent-electron system. All this comes at the expense of having to find reasonable approximations to the exact energy functional, whose exact form remains, excluding some special low-dimensional model systems³⁸, unknown. Unfortunately, only the HOMO eigenvalue (negative IP) would be correctly described in exact KS-DFT, but even for this case the most common and computationally inexpensive

approximations that are employed provide values that are in poor correspondence with the first particle-removal energies. Within the local-density approximation (LDA)³⁹ or the generalized gradient approximations (GGAs) [with a common example as Perdew-Burke-Ernzerhof (PBE) approximation⁴⁰], the HOMO values are systematically too high in energy (underestimating the IP), in part due to large self-interaction errors (SIE)³⁹. Self-interaction is not only responsible for incorrect electron binding energies, but also for the spatial over-delocalization of charge densities and KS orbital wave functions^{33,41}, which in turn affects multiple aspects of ground-state predictions, including molecular equilibrium bond lengths, molecular dissociation energies, adsorption and transition states energies⁴².

1. Piecewise linearity and self-interaction errors

With the purpose of correcting SIE, Perdew and Zunger³⁹ introduced an orbital-dependent density (ODD) functional, termed here PZ, obtained by subtracting from an approximate functional E^{app} the sum of the Hartree ($E_{\text{H}}[\rho_i]$) and exchange-correlation ($E_{\text{xc}}[\rho_i]$) energies of each single (fully or partially) filled orbital $\rho_i(\mathbf{r})$:

$$E^{\text{PZ}} = E^{\text{app}} - \sum_i (E_{\text{H}}[\rho_i] + E_{\text{xc}}[\rho_i]). \quad (1)$$

While leading to an exact formulation for one-electron systems, the PZ functional generally overcorrects the SIE⁴² in many-electron systems, resulting in incorrect ionization energies, structural predictions, and reaction paths^{42,43}.

As a route towards a good description of the properties of many-electron systems, several authors^{33,35,41,44} have suggested a new definition of the SIE based on the lack of PWL of the total energy as a function of the (fractional) number of electrons. Indeed, it can be shown that a dependence of $E(N)$ on the particle number N that is convex tends to delocalize total and orbital densities, whereas functionals for which $E(N)$ is concave (such as PZ) lead to over-localization⁴⁵.

As mentioned, KC functionals^{26–32} can be seen as a generalization of DFT+U aimed at explicitly enforcing PWL to an entire electronic manifold. These functionals are obtained by removing, orbital-by-orbital, the non-linear (Slater) contribution to the total energy and by replacing it by a linear (Koopmans) term. This linear term is chosen either following Slater’s suggestion⁴⁶, i.e., proportional to the orbital energy at half orbital filling (in which case the KC functional is simply labeled K^{28,32}), or as the difference between the energies of the two adjacent electronic configurations with integer occupation; this latter is labeled KI (“I” standing for “integral”). The numerical differences between these two approximations are largely negligible, and we focus here on KI, which is simpler to implement. KI functionals, described in detail

in Ref. 32, are obtained from an approximate functional E^{app} as

$$E^{\text{KI}} = E^{\text{app}} + \sum_i \alpha_i \Pi_i^{\text{KI}}, \quad (2)$$

where

$$\begin{aligned} \Pi_i^{\text{KI}} = & -E_{\text{Hxc}}[\rho] + E_{\text{Hxc}}[\rho - \rho_i] \\ & + f_i(-E_{\text{Hxc}}[\rho - \rho_i] + E_{\text{Hxc}}[\rho - \rho_i + n_i]). \end{aligned} \quad (3)$$

In the above equations, $E_{\text{Hxc}} = E_{\text{H}} + E_{\text{xc}}$, $n_i(\mathbf{r}) = |\varphi_i(\mathbf{r})|^2$, $\rho_i(\mathbf{r}) = f_i|\varphi_i(\mathbf{r})|^2$, $\rho(\mathbf{r}) = \sum_i f_i|\varphi_i(\mathbf{r})|^2$, and α_i are orbital-dependent screening coefficients that account for orbital relaxations, since for $\alpha_i = 1$, the KI functional described in Eq. (2) fulfills exactly the generalized Koopmans condition at frozen orbitals²⁹. The α_i coefficients, which can be computed from first principles (see below), are generally smaller than one. It should be noted that the KI functional is piecewise linear with respect to fractional changes in the particle number, but it does not change the total energy nor the ground-state wave function (and consequently the one-body density matrix) of the approximate "base" functional whenever the system has an integer number of particles. We also note in passing that, using this terminology, the ensemble-DFT correction of Ref. 44 is equivalent to the KI functional, when applied to the frontier orbitals.

It is possible to modify the definition in Eq. (2) so as to obtain a functional which is, similarly to PZ, exact in the one-electron limit and variational, while remaining approximately self-interaction free in the many-electron case (we stress that the generalized Koopmans condition is stronger than just being many-electron self-interaction free, since it applies simultaneously to all orbitals). Such functional can be obtained by applying a KI correction on top of the PZ functional, resulting in the following definition:

$$E^{\text{KIPZ}} = E^{\text{app}} + \sum_i \alpha_i \{ \Pi_i^{\text{KI}} - f_i E_{\text{Hxc}}[n_i] \}. \quad (4)$$

The Koopmans orbital-by-orbital linearity condition imposed through Eq. (2) or (4) leads to an ODD formulation in which the energy functional depends on the density of the individual orbitals. As such, differently from DFT functionals, but similarly to other ODD functionals such as PZ, KC functionals are not invariant under unitary transformations within the manifold of filled orbitals^{22,32,47–49} and the *variational orbitals* $|\varphi_i\rangle$ that minimize the functional are different from the eigenstates or *canonical orbitals* $|\phi_m\rangle$ that diagonalize the matrix of Lagrange multipliers, as discussed, e.g., in Refs. 47–49. The strategy that we use to minimize KC functionals, which follows the ensemble-DFT algorithm⁵⁰ (note that this is unrelated to the ensemble-DFT correction of Ref. 44) for the case of orbital-density-dependent functionals, consists of two nested steps: (i) a minimization with respect to unitary transformations at fixed orbital

manifold (inner loop), that leads to a projected, unitary-covariant functional of the orbitals only enforcing the Pederson condition⁵¹; (ii) a variational optimization of the orbital manifold of this projected functional⁴⁹ (outer loop).

The generalized eigenvalue equation obtained within the ODD formalism reads

$$\hat{H}^{\text{app}}|\phi_m\rangle + \hat{\Sigma}_m|\phi_m\rangle = \varepsilon_m|\phi_m\rangle \quad (5)$$

with

$$\Sigma_m(\mathbf{r})\phi_m(\mathbf{r}) = \sum_i v_i^{\text{ODD}}(\mathbf{r})\varphi_i(\mathbf{r})U_{im}^\dagger \quad (6)$$

where the unitary matrix U transforms the variational orbitals $|\varphi_i\rangle$ into the canonical ones $|\phi_m\rangle = \sum_i |\varphi_i\rangle U_{im}^\dagger$, and the ODD potential is given by

$$v_i^{\text{ODD}}(\mathbf{r}) = \sum_j \frac{\delta \Pi_j^{\text{ODD}}}{\delta \rho_i(\mathbf{r})}. \quad (7)$$

Full details about the expressions for $v_i^{\text{ODD}}(\mathbf{r})$ when the ODD correction is KI or KIPZ are given in Ref. 32. As discussed in Refs. 22 and 36, ODD canonical orbitals can be interpreted as approximations to Dyson orbitals (solutions of quasiparticle equations), and their energies as particle-removal energies; this interpretation will be supported by the results of the present work.

2. Calculation of the screening coefficient

In the definitions of KI and KIPZ functionals, the multiplicative factors α_i are meant to account for orbital relaxation²⁹ and should be orbital dependent. Often, we simply choose them to be all equal to a single (effective) α value chosen so that the IP of a neutral molecule is equal to EA of the molecular cation^{28,52}. The optimal value of α is computed through the secant recursion method:

$$\alpha_{n+1} = \alpha_n + \frac{(1 - \alpha_n)(\epsilon_{L,N-1}^{\alpha_n} - \epsilon_{H,N}^{\alpha_n})}{(\epsilon_{L,N-1}^{\alpha_n} - \epsilon_{H,N}^{\alpha_n}) - (\epsilon_{L,N-1}^{\alpha_{n-1}} - \epsilon_{H,N}^{\alpha_{n-1}})} \quad (8)$$

where $\epsilon_{L,N-1}^{\alpha_n}$ and $\epsilon_{H,N}^{\alpha_n}$ are the LUMO energy of the positively ionized system and the HOMO energy of the neutral system, respectively. As already mentioned, the procedure to determine α is fully *ab initio*, since all HOMO and LUMO eigenvalues appearing in Eq. (8) are the result of numerical simulations involving an approximate DFT functional with Koopmans' corrections. As shown by Dabo et al.²⁸ and Borghi et al.³², a constant screening coefficient is sufficient to accurately predict IPs and EAs of a variety of atomic and molecular systems, even though an orbital-dependent α might be more accurate in the case of large or extended systems, or in systems composed of sub-systems having very different orbital relaxation properties.

Table I. Values of $\alpha^{(v)}$ and $\alpha^{(c)}$ computed on top of the initial structures, PBE (@PBE), or on top of the structure resulting from scf-KIPZ optimization (@KIPZ) of different DNA/RNA nucleobases.

	$\alpha^{(v)}$ @PBE	$\alpha^{(v)}$ @KIPZ	$\alpha^{(c)}$ @PBE	$\alpha^{(c)}$ @KIPZ
A	0.4523	0.4713	0.9822	0.9934
T	0.4654	0.4732	1.0324	1.0184
U	0.4801	0.4886	0.9422	0.9681
C ₁	0.4389	0.4560	0.9725	0.9873
G ₂	0.4225	0.4465	0.9352	0.9523

In fact, when computing EAs of molecules, results improve³⁶ when choosing as screening coefficient the one appropriate for the LUMO, which can be obtained easily from Eq. (8) replacing N with $N + 1$. This “anionic” screening, which we will refer to as $\alpha^{(c)}$, where the letter “c” stands for “conduction” (as opposed to the “neutral” or “valence” one, for which we will use the symbol $\alpha^{(v)}$) leads to much better results for the binding energies of empty orbitals. In Sec. IIIB and Table II, we will show that this choice reproduces correctly the energies and the orbital ordering of the empty states of nucleobases. Meanwhile, we show in Table I the values of $\alpha^{(v)}$ and $\alpha^{(c)}$ computed from Eq. (8) for the most stable nucleobases. One can immediately remark that $\alpha^{(c)}$ is very close to 1 (except for the case of the T molecule which has $\alpha^{(c)} > 1$, meaning that the LUMO state is more delocalized compared to the other orbitals), suggesting that the orbital relaxation that takes place in the anionic system upon removal of its extra electron is very small.

3. Geometry and screening optimization

In previous work^{28,29,31,32,36} the calculation of $\alpha^{(v)}$ was performed at fixed geometry, assuming the variation of the screening during the geometry optimization to be negligible. In this work we explore self-consistency for $\alpha^{(v)}$ with respect to molecular geometry. The workflow is presented in Fig. 1, and starts with an initial guess that can be chosen from the geometries computed from standard DFT. Then, the value $\alpha^{(v)}$ corresponding to this geometry is estimated according to Eq. (8).

After that, a geometry optimization is performed with the KIPZ functional, using the computed $\alpha^{(v)}$ as screening coefficient. A new value of $\alpha^{(v)}$ is then calculated for the new geometry. This completes a self-consistent loop which is iterated until the inter-atomic force (F) and the change in screening coefficient ($\Delta\alpha$) between subsequent steps are smaller than a given threshold. In Table I, we show the change of the α -coefficients during the self-consistent procedure. The detailed results obtained with this optimization scheme are reported in Sec. IIID, where we discuss the accuracy of Koopmans-compliant functionals in predicting the geometry of DNA and RNA bases.

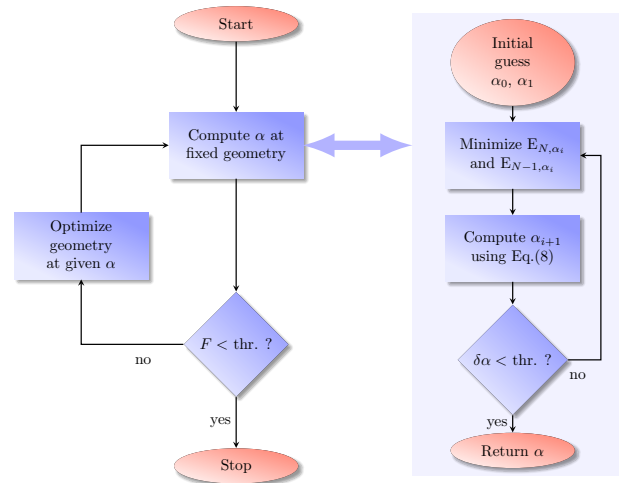


Figure 1. Diagram of the self-consistent KIPZ scheme used to optimize the screening factor α ($\alpha^{(v)}$) along with molecular structure.

III. RESULTS AND DISCUSSION

A. Technical details

In this Section we present the results obtained by applying the above scf-KIPZ procedure to compute binding energies, ultraviolet photoemission spectra, and optimized geometries for all nucleobases. Since our code (a modified version of cp.x from the Quantum-ESPRESSO distribution⁵³) works with periodic-boundary conditions, we place the molecules inside an orthorhombic cell with at least 18 Bohr of vacuum on each side, and we add reciprocal-space counter-charge corrections⁵⁴ to the electrostatic energies and potentials in order to suppress the Coulomb interactions between periodic replicas. All calculations are performed using norm-conserving pseudopotentials⁵⁵ and the cutoff for the plane wave expansion of wavefunctions is set to 60 Ry. Moreover, in all PZ and KIPZ calculations presented in this paper, the orbital-density dependent corrections (either PZ or KIPZ) are computed on top of the PBE functional. In previous work, we showed that the combination of PZ and KIPZ with PBE as base functional produces the best results for the electronic eigenvalue spectrum when minimized on the Hilbert space of complex wave functions^{32,56}, which is the procedure adopted here.

The set of all possible nucleobase tautomers that could be the subject of our study is quite large; in this work we choose to study only the most stable ones, as reported in Fig. 2. From our DFT calculations we find that there is a single tautomer that is energetically favored for A, T, and U (for these bases the total energy difference between the less stable tautomers and the most stable one is > 0.43 eV). This fact yields the normalized Boltzmann weighting factors evaluated at room temperature (300 K), computed using the total energies and neglecting vi-

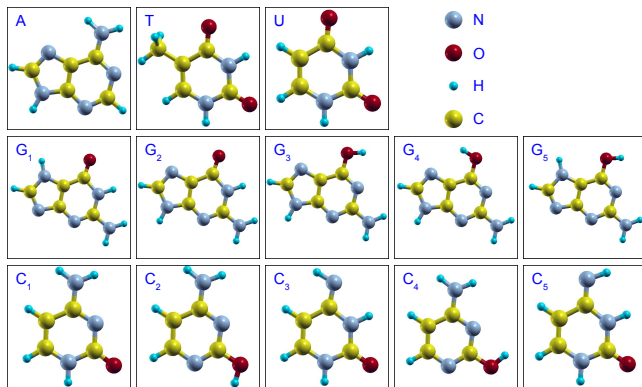


Figure 2. Atomic structure of A, T, U, and of the five most stable tautomers of G and C considered in this study.

brational entropy, for the most stable tautomers are one and the others are zero). Instead, for the C and G nucleobases, we find five different tautomeric forms with very close total energies, and we include all of them in our set. For each nucleobase, we compute the KIPZ total energy of different tautomers using only one value of α , which, for C and G, is computed as the average of the screening coefficients of the 5 tautomers. The normalized Boltzmann weighting factors for the five tautomers of C and G are $C_1 : C_2 : C_3 : C_4 : C_5 = 0.32 : 0.47 : 0.12 : 0.06 : 0.03$ and $G_1 : G_2 : G_3 : G_4 : G_5 = 0.44 : 0.20 : 0.18 : 0.16 : 0.02$, respectively. These ratios are in reasonable agreement (correct order) with coupled-cluster CCSD(T) results^{57,58}, giving $C_1 : C_2 : C_3 = 0.23 : 0.64 : 0.13$ and $G_1 : G_2 : G_3 = 0.60 : 0.27 : 0.13$. In Sec. III C we show how the experimental UPS of C and G can be reproduced with good accuracy from the average of the spectra of different tautomers using the KIPZ weighting factors.

B. Binding energies of frontier orbitals

In this section we assess the accuracy of the KIPZ functional in computing vertical IPs and EAs, and more in general in predicting the binding energies and the molecular orbital (MO) character of the frontier occupied and unoccupied orbitals.

In Fig. 3 we present the KIPZ predictions for the first-IP of DNA and RNA bases as compared to experiments, together with results computed at the DFT-PBE level and using the Slater Δ SCF method. The latter is often considered an accurate approach for computing the first IP and EA for small, finite systems. Experimental data resolved for each C and G tautomer cannot be obtained, due to the closeness in energy of the tautomers (see the discussion at the end of Sec. III A); we therefore use the same value as a reference for all of them.

From the data in Fig. 3 we find that within PBE the value of the IP is underestimated with a mean absolute error (MAE) of about 3.10 eV with respect to exper-

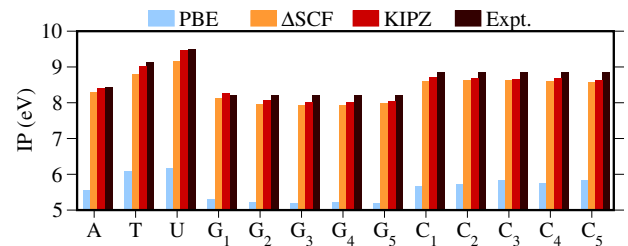


Figure 3. Vertical IPs computed with the standard DFT (PBE exchange-correlation functional), Δ SCF, and KIPZ, for all the nucleobases and their tautomers considered in this work. The experimental numbers used as a reference for vertical ionization potentials are the average values computed from the various experiments listed in Table II.

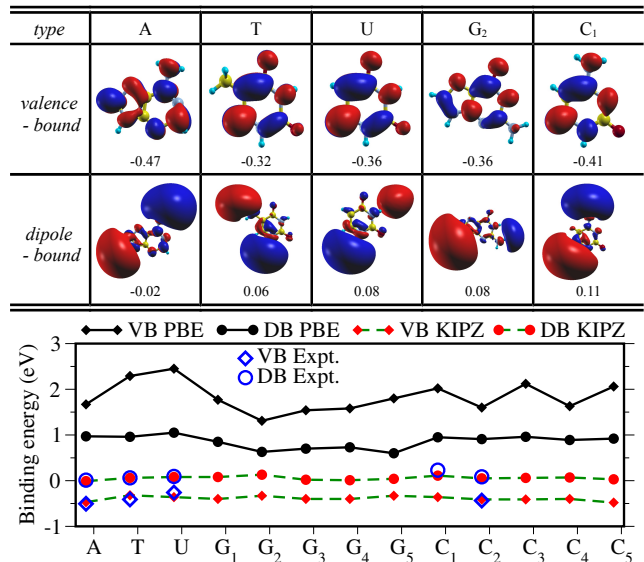


Figure 4. (Top panel) density isosurfaces for valence-bound (VB) and dipole-bound (DB) unoccupied states of the five DNA nucleobases shown in Table II. All are computed with the KIPZ functional. The two different colors refer to the sign of the wave functions. The values listed below the plots are the corresponding orbital binding energies. (Bottom panel) Orbital binding energies of VB and DB orbitals computed with the standard DFT (PBE exchange-correlation functional) and KIPZ for the nucleobases considered in this work. The experimental numbers used as a reference are taken from the experimental EA energies listed in Table II.

iments. The incorrect predictions are cured by KIPZ, which reduces the error down to 0.1 eV. The accuracy of KIPZ compares favorably to that of the Slater Δ SCF method, which has a MAE of ~ 0.15 eV. We also recall here that all the theoretical predictions for PBE-, PZ-, GW- and KIPZ- IPs are obtained from the negative of the orbital energy of the HOMO.

In Table II we show a comparison between KIPZ predictions for frontier orbital binding energies (i.e. HOMO-1, HOMO, LUMO and LUMO+1) and those of other

Table II. Binding energies and molecular orbital characters of HOMO-1 (σ), HOMO (π), LUMO (dipole-bound [DB] state) and LUMO+1 (valence-bound [VB] state) canonical orbitals obtained from KIPZ eigenvalues in comparison with standard DFT (PBE exchange-correlation functional), PZ, G_0W_0 , scf-GW and quantum chemistry calculations (CASPT2: complete active space with second-order perturbation theory; CCSD(T): coupled-cluster with singles, doubles, and perturbative triple excitations) and experimental data. Experimental values are taken as a reference for the calculation of mean absolute errors (MAE). Orbital energies within KIPZ are obtained with the $\alpha^{(v)}$ screening factor for “valence” filled orbitals, and with $\alpha^{(v)}$ for empty “conduction” orbitals. For the empty orbitals we report, in parentheses, orbital energies obtained using the $\alpha^{(v)}$ screening factor instead.

	KIPZ orb.	PBE	PZ	G_0W_0 (PBE) ^a	scf-GW ^b	CASPT2/CCSD(T) ^{c,d,e}	KIPZ	Experiment ^{f,g,h,i,j,k,l,m}
A	[L+1]VB	1.67	1.76	-0.25	-1.14	-0.91 ^c	(-0.32)-0.47	-0.56~-0.45 ^g
	[L]DB	0.72	0.63	-0.31			(0.13)-0.02	0.012 ^m
	[H] π	5.55	9.90	7.99	8.22	8.37 ^d /8.40 ^d	8.41	8.3~8.5 ^f /8.47 ^h
	[H-1] σ	5.82	10.81	8.80	9.47	9.05 ^d	9.01	9.45 ^h
T	[L+1]VB	2.29	2.49	0.24	-0.67	-0.60 ^c /-0.65 ^c	(0.18)-0.32	-0.53 ~ -0.29 ^g
	[L]DB	0.59	0.92	-0.26			(0.33)0.06	0.062 ~ 0.068 ^l
	[H] π	6.08	10.92	8.63	9.05	9.07 ^d /9.04 ^d	9.02	9.0 ~ 9.2 ^f /9.19 ^h
	[H-1] σ	6.11	12.25	8.94	10.41	9.81 ^d	9.77	9.95 ~ 10.05 ^f /10.14 ^h
U	[L+1]VB	2.44	2.64	0.23	-0.64	-0.61 ^c /-0.64 ^c	(0.26)-0.36	-0.3~-0.22 ^g
	[L]DB	0.82	0.98	-0.29			(0.38)0.08	0.093 ^l
	[H] π	6.17	11.22	8.99	9.47	9.42 ^d /9.43 ^d	9.45	9.4~9.6 ^f
	[H-1] σ	6.4	12.36	9.07	10.54	9.83 ^d	9.97	10.02 ~ 10.13 ^f
G ₂	[L+1]VB	1.29	1.53	-0.43	-1.58	-1.14 ^c	(-0.17)-0.36	
	[L]DB	0.63	0.67		-0.20	0.056 ~ 0.065 ^e	(0.46)0.08	
	[H] π	5.22	9.62	7.64	7.81	8.09 ^d /8.09 ^d	8.07	8.0 ~ 8.30 ^f /8.30 ⁱ /8.26 ^j
	[H-1] σ	5.86	11.56	8.67	9.82	9.56 ^d	9.25	9.90 ⁱ /9.81 ^j
C ₁	[L+1]VB	2.01	2.19	-0.02	-0.91	-0.69 ^c /-0.79 ^c	(-0.03)-0.41	-0.55~-0.32 ^g
	[L]DB	0.62	0.90	-0.23			(0.38)0.11	0.23 ^k
	[H] π	5.70	10.50	8.18	8.73	8.73 ^d /8.76 ^d	8.70	8.80~8.90 ^f /8.89 ^h
	[H-1] σ	6.26	10.96	8.5	9.89	9.42 ^d	9.12	9.45 ⁱ /9.55 ^h
MAE	[L+1]VB	2.51	2.67	0.45	0.44	0.32	(0.43)0.06	
	[L]DB	0.59	0.76	0.29			(0.21)0.05	
	[H] π	3.08	1.61	0.54	0.17	0.08	0.09	
	[H-1] σ	3.70	1.80	0.99	0.25	0.25	0.36	

^a Reference 11

^b Reference 59

^c Reference 10

^d Reference 9

^e Reference 60

^f Collected in reference 9

^g Collected in reference 10

^h Reference 5

ⁱ Reference 61

^j Reference 7

^k Reference 16

^l Reference 62

^m Reference 63

theoretical approaches, including experimental values when available. The accuracy of KIPZ is found higher than that of PZ, G_0W_0 , and self-consistent GW (scf-GW). We note that KIPZ compares favorably also with highly-accurate quantum chemistry methods such as complete active space with second-order perturbation theory (CASPT2) and coupled-cluster with singles, doubles, and perturbative triple excitations [CCSD(T)]. In Table II we report also the binding energies of the first two empty orbitals, recalling again that we take the binding energy of the LUMO orbital as its vertical EA.

Besides the experimental difficulties, the theoretical

calculation of nucleobase EAs is also non-trivial, mostly because of the large polarity of DNA bases (dipoles larger than 2.5 Debye). Such polarity allows for the existence of a stable, but very weakly bound, “dipole-bound” (DB) anionic state. This state is close in energy to an anionic valence bound (VB) state, characterized by an extra electron occupying a valence anti-bonding molecular orbital. The nature of these two different states can be investigated using different experimental techniques. The energy of DB states, which are very weakly bound but stable, can be measured from negative ion photoelectron spectroscopy⁶². VB states instead are accessi-

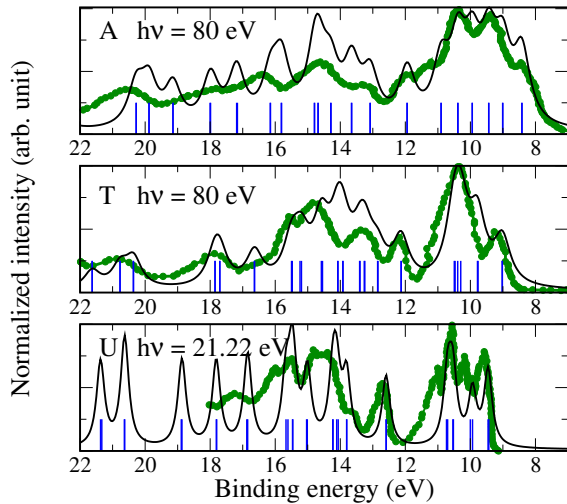


Figure 5. UPS spectra for adenine, thymine and uracil molecules calculated using the KIPZ functional (black line), and plotted as a function of electron binding energy ($h\nu - \hbar^2 k^2 / (2m)$). For adenine and thymine the calculations are carried out for an incoming photon energy of 80 eV, and compared with experimental gas-phase UPS measurements (green dots) using radiation with the same photon energy⁵. For uracil both experimental and simulation data are for a 21.22 eV⁶⁴ photon energy. The blue bars in each plot mark the energies of KIPZ canonical orbitals.

ble through electron transmission spectroscopy⁶⁵. With this technique it was proved that adding an electron to a VB state requires a positive energy, which means that the electronic configuration of a nucleobase with an extra electron on its anti-bonding orbital is unstable. When it comes to numerical simulations, the instability of a state can be established from the negative sign of its binding energy. Interestingly, not all theoretical methods are able to access both types (VB and DB) of anionic states. Accurate CCSD(T) simulations¹⁰ on neutral and negatively charged DNA nucleobases, for instance, result in negative EAs, and predict the anionic state to be of the VB type. This disagrees with experimental findings which suggest the existence of a weakly-bound DB state. The inability of these CCSD(T) calculations to predict the existence of a frontier DB state has been questioned¹⁰ and connected to the basis sets used in the simulations, unable to describe orbitals with a highly diffuse character in the vacuum region around the molecule. The same issue affects GW calculations when they are performed with localized basis sets⁵⁹, while it does not apply to plane-wave-based GW methods. Indeed Qian, Umari and Marzari¹¹ were able to obtain an empty VB state with negative energy (positive binding energy, thus able to bind an extra electron) in the spectrum of neutral guanine.

Using our plane-wave based code we are able to reproduce both VB and DB states, and their orbital densities and energies are shown in Fig. 4. By performing PBE and KIPZ calculations on neutral nucleobases, we can access the orbital energies of empty states and as-

sess their ability to bind extra electrons to the molecule. PBE and KIPZ orbital energies for VB and DB empty states are shown in Fig. 4. PBE results not only deviate dramatically from experiment, but reverse the order of VB and DB energies, and predict both states to be able to bind electrons. The addition of Koopmans’ corrections reverses the order of orbital energies and pushes the binding energies of VB states to negative values, restoring the agreement with experimental results⁶⁵, showing VB states to be unstable. The agreement of our results for guanine with the G_0W_0 results of Qian *et al.*¹¹ is a further proof of the reliability of the KIPZ functional. In Table II we show the binding energies of frontier orbitals compared to experiments and different theoretical results. It is worth to stress again that for an accurate prediction of the binding energies of empty orbitals we use as screening coefficient not the screening coefficient of the neutral system ($\alpha^{(v)}$), but the one of the anionic system ($\alpha^{(c)}$).

C. Ultraviolet photoemission spectroscopy

The capability of the KIPZ functional of predicting not only the binding energies of frontier orbitals, but also those of deeper states, makes it a promising tool for the calculation of photoemission spectra. In this section we show how KIPZ can describe the position of (quasiparticle) photoemission peaks as well as their strengths and shapes. We obtain theoretical photoemission spectra following the well established three-step model within the sudden approximation⁶⁶. This approach treats the photoexcitation as a transition from an electronic initial state $|\Phi_0^N\rangle$ — which is the ground state with energy E_0^N — into an excited N-particle state $|\Phi_{i,\mathbf{k}}^N\rangle = |\Phi_i^{N-1}\rangle \otimes |\xi_{\mathbf{k}}\rangle$ of energy $E_{i,\mathbf{k}}^N$, built from the i^{th} excited state of the singly ionized system (with energy E_i^{N-1}) and the wave function $\xi_{\mathbf{k}}$ of the ejected electron, approximated by a plane wave with wave vector \mathbf{k} . Here, it is worth stressing that $\xi_{\mathbf{k}}$ can be approximated in different ways, the simplest one being a plane wave, which can be further orthogonalized to the initial states (with the aim of improving the description of the final states). However, the discrepancy between these approaches has been shown to emerge mainly when looking at states with large binding energies⁶⁷ (low kinetic energy photoelectrons), so in the present work we adopt the plane wave approximation, which is expected to be accurate for high kinetic energy photoelectrons. The total photoemission intensity can be described, to first order in perturbation theory, through the Fermi’s golden rule⁶⁶ as

$$I^{(\nu)} \propto \sum_{i,\mathbf{k}} |\langle \Phi_0^N | \mathbf{A} \cdot \sum_j \hat{\mathbf{p}}_j | \Phi_{i,\mathbf{k}}^N \rangle|^2 \delta(h\nu + E_0^N - E_{i,\mathbf{k}}^N), \quad (9)$$

which contains the squared modulus of the light-matter interaction operator in the dipole approximation — where \mathbf{A} is the amplitude of the semi-classical vector-

potential and $\hat{\mathbf{p}}_j = -i\hbar\nabla_j$ is the linear momentum operator for the j^{th} electron, Equation (9) can be written in terms of single-particle Dyson orbitals $\phi_i^{\text{d}}(\mathbf{r}) = \langle \Phi_i^{N-1} | \hat{\Psi}(\mathbf{r}) | \Phi_0^N \rangle$ and binding energies $E_i^{\text{b}} = E_i^{N-1} - E_0^N$, as⁶⁸:

$$I^{(\nu)} \propto \sum_{i,\mathbf{k}} |\langle \phi_i^{\text{d}} | \mathbf{A} \cdot \hat{\mathbf{p}}_i | \xi_{\mathbf{k}} \rangle|^2 \delta \left(h\nu - E_i^{\text{b}} - \frac{\hbar^2 \mathbf{k}^2}{2m} \right). \quad (10)$$

More details on the calculation of $I^{(\nu)}$ can be found in Ref. 36. The excitation energy is now expressed in terms of the kinetic energy $\hbar^2 \mathbf{k}^2 / 2m$ of the ejected electron and its binding energy E_i^{b} defined as the negative of the Dyson orbital energy ε_i^{d} .

Dyson orbitals, whose energies are the poles of the one-body Green's function, fulfil (at least for discrete states¹⁸) the quasiparticle equation

$$[\hat{T} + \hat{v} + \hat{\Sigma}(\varepsilon_i^{\text{d}})] |\phi_i^{\text{d}}\rangle = \varepsilon_i^{\text{d}} |\phi_i^{\text{d}}\rangle, \quad (11)$$

where \hat{v} is the sum of the external and Hartree potentials and $\hat{\Sigma}$ is the electron-electron self-energy. The calculation of Dyson orbitals should in principle be carried out within the framework of many-body perturbation methods¹⁸. However, there is a strong analogy between Eq. (11) and the generalized eigenvalue equation [Eq. (6)] of Koopmans-compliant functionals, where the local and orbital density dependent operator \hat{v}_i acting on the variational orbitals $|\varphi_i\rangle$ can be seen as a simplification of a non-local and frequency-dependent self-energy, as argued in Ref. [22]. The canonical orbitals produced by an orbital-density-dependent calculation have in this perspective a natural interpretation as Dyson orbitals, and their energies as particle-removal energies. This justifies the choice to use them to construct photoemission spectra, in a framework which is computationally much less expensive than solving Eq.(11) within state-of-the-art Green's function methods.

The KIPZ results for nucleobases' spectra are shown in Figs. 5, 6 and 7. Theoretical spectra are compared directly with experimental data only for the bases that have one single stable tautomer. This is not the case for G and C, for which the comparison is carried out between the experimental value and the Boltzmann average of the spectra of the five different tautomers. All photoemission spectra are computed with incoming photon energies $h\nu$ taken from the referenced experiments. For A, T, and C $h\nu = 80$ eV (Ref. 5), for G $h\nu = 100$ eV (Ref. 7), and for U $h\nu = 21.22$ eV (Ref. 64). A detailed study of the effects of different photon energies on the shape and intensities of the photoemission spectra is shown in Figs. S1 and S2 in the Supporting Information⁶⁹.

The agreement between computed spectra and experiments is remarkable, not only for the peak positions but also for the shapes and intensities of the spectral peaks. The success of KIPZ can be explained through its ability to correct KS eigenvalues of approximate DFT by aligning them to particle removal energies through the Koopmans' condition, and by inheriting from PZ the property

of being exact in the one-electron limit³². Such behavior is essential for the prediction of fundamental gaps and excitation energies⁷⁰. Another feature of KIPZ is its ability to modify not only the electronic excitation energies of approximate DFT, but also the manifold of electronic orbitals (i.e., the single-particle density-matrix)³². A change in the shape of single-particle orbitals, which results in a change of the one-body density-matrix of quasiparticles, affects both photoemission peak intensities and positions, ultimately affecting the accuracy of the simulated spectra.

The availability of experimental data from angle-resolved photoemission spectroscopy (ARPES) for DNA and RNA nucleobases enables us to comment also on the ability of KIPZ to predict the correct ordering, as well as the binding energy, of the orbitals close to the HOMO. The KIPZ results show that the type and order of the five orbitals with the lowest binding energies in the stable tautomers of A, T and U are the same, i.e., from HOMO to HOMO-4 the character of the orbitals is $\pi, \sigma, \pi, \sigma, \pi$ (see Fig. S3 in the Supporting Information⁶⁹ for a plot of the orbital densities in each of them), in agreement with ARPES measurements and with other highly-accurate quantum chemistry calculations (using the full third-order algebraic diagrammatic construction [ACD(3)]⁵ or the equation-of-motion coupled-cluster with single and double substitutions [EOM-IP-CCSD]¹⁷).

In the case of C and G, orbital ordering depends on the tautomer considered. The KIPZ orbital ordering which matches experimental ARPES measurements is the one of the most stable tautomer, i.e., the orbital ordering of C₂ ($\pi, \sigma, \pi, \sigma, \pi$) and G₁, respectively. Other tautomers such as C₁, C₃ and C₅ have an orbital ordering in which the 2nd and 3rd orbitals are swapped with respect to C₂, while the ordering of G₁, G₃, G₄ and G₅ differs from G₂ by a swap of the 3rd and 4th orbitals. More details on the nature of the orbitals close to the HOMO can be found in Figs. S4 and S5 in the Supporting Information⁶⁹.

It is worth to stressing that while for C and G the orbital character (σ or π) and ordering appear to be determined from the most stable tautomers, the prediction of shape, intensity, and peak positions in the photoemission spectra requires the contribution of all tautomers, each weighted with its Boltzmann factor. If we look in detail at the spectrum of each single tautomer of C or G (Figs. 6 and 7), we see that from their overall appearance these can be divided into two groups. In the case of C, the spectra of C₁, C₃ and C₅ are similar, and are characterized by a low-energy prominent peak which appears in the Boltzmann-averaged spectrum on the first panel of Fig. 6. We can thus say that the electronic excitations of C₁, C₃ and C₅ are fundamental in determining the experimental spectrum of C, even though they do not include the most stable tautomer (C₂). Concerning C, we can determine a second group of tautomers with similar spectra, i.e., C₂ and C₃. The two groups of tautomers differ by the fact that a hydrogen atom is bound to an oxygen in the case of C₂ and C₃, and to a nitrogen in the case of

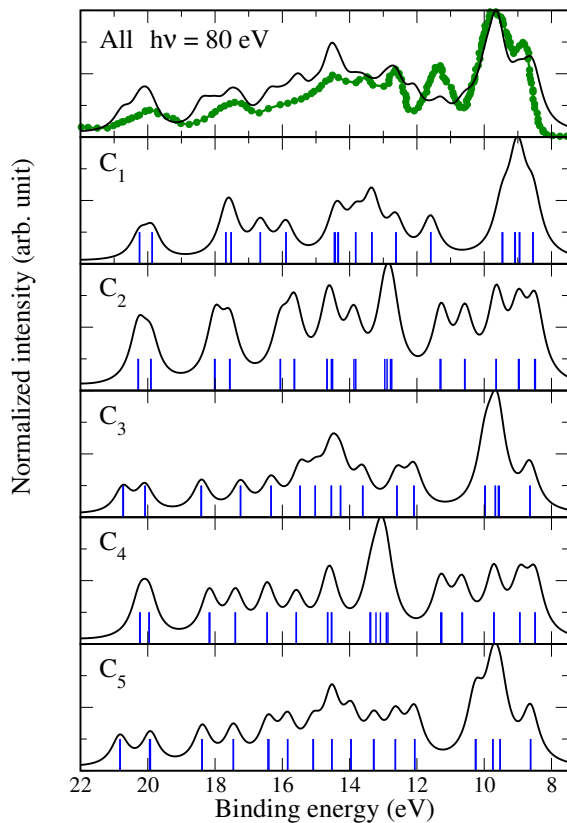


Figure 6. UPS spectra for cytosine tautomers calculated with the KIPZ functional (black line), and plotted as a function of electron binding energy ($h\nu - \hbar^2 k^2 / (2m)$). The calculations are performed for an incoming photon energy of 80 eV. In the top panel we show the total theoretical spectrum (black line), which is averaged over the five most stable tautomers with Boltzmann weighing factors, and compared with experimental gas-phase UPS measurements (green dots) using radiation with the same photon energy⁵. The blue bars in each plot mark the energies of KIPZ canonical orbitals.

C_1 , C_3 and C_5 (see also Fig. 2). Correspondingly, one can observe similarities, which are less pronounced than in the case of C , in the spectra of G_1 and G_2 . These two first tautomers can be grouped together and contrasted to G_3 , G_4 and G_5 , which have in common the presence of a hydrogen atom bound to an oxygen rather than a nitrogen. All the above remarks emphasize how photoemission peaks and the overall photoemission spectra can in principle be used in order to extract information not only on the type of nucleobase, but also on the precise tautomeric forms present in a sample.

D. Geometry optimization

In this section we discuss the effects of structural optimization within the KIPZ framework on the electronic properties of DNA and RNA nucleobases. The self-consistent optimization procedure, involving atomic posi-

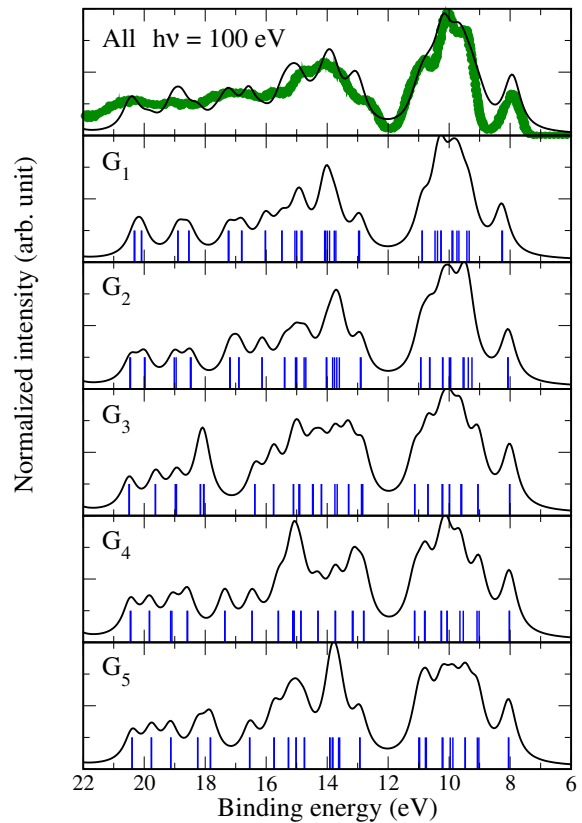


Figure 7. Same as Fig. 6 for guanine, and for an incoming photon energy of 100 eV. Experimental data are from Ref. 7.

tions and screening coefficient α , was outlined in Sec. II 3 and Fig. 1. In Table III, we test accuracy in computing intra-molecular bond lengths and angles by comparing the mean absolute error (MAE) with respect to experimental data of the structural predictions coming from several *ab initio* methods. In Table III we show data obtained by PBE, PZ, KIPZ, and PBE0.

Experimental data are taken from the Cambridge Structural Database by Clowney et al.⁷¹ where bond lengths are measured at room temperature through high-resolution X-ray and neutron diffraction; these data should be compared to the most stable tautomeric form of each nucleobase. In order to be consistent with these data, in Table III we show the MAE with respect to experiment only for bond lengths and angles of those bases that have a single stable tautomer at room temperature, namely A, T and U. More detailed information about structural properties is available in Table S1, S2 and S3 in the Supporting Information⁶⁹.

By looking at Table III, one sees that the maximum error in predicting structural properties is displayed by DFT-PBE (MAE > 1.39%). The relatively smaller accuracy of this functional can be partly explained from its self-interaction error, which results in a slightly increased spread of orbital densities. This can in turn affect geometrical properties, and in molecules it typically results in bond-lengths that exceed experimental values³².

Table III. Relative mean absolute error of bond lengths and angles of A, T, and U molecules computed using PBE, PZ, KIPZ, and PBE0 methods, compared with experiments⁷¹.

% error		PBE	PZ	KIPZ	PBE0
\bar{R}_{ij}	A	1.39	0.67	0.70	0.56
	T	1.77	0.89	0.55	0.85
	U	1.76	0.96	0.70	0.93
$\bar{\theta}_{ijk}$	A	0.51	0.52	0.56	0.54
	T	0.74	0.59	0.64	0.63
	U	1.05	0.69	0.64	0.81

In contrast, the PZ functional, usually over-correcting the self-interaction error in molecules, tends to underestimate bond lengths. The KIPZ functional, thanks to its *ab initio* screening factor, can interpolate between the two opposite behaviors of PBE and PZ, resulting in more accurate estimates of geometrical parameters. Our results show that the accuracy of KIPZ (MAE $\sim 0.6\%$) in computing bond lengths is better than PBE, and close to that of the PBE0 hybrid functional. The same cannot be said for angles, for which the discrepancy between the different methods shown is much smaller, and shows no clear trend.

Table IV. Absolute dihedral angles (β_1 and β_2) of the nucleobases with amino group. For A $\beta_1 = \angle \text{H}_2\text{N}_{10}\text{C}_4\text{N}_3$ and $\beta_2 = \angle \text{H}_1\text{N}_{10}\text{C}_4\text{C}_5$; for G_{1..5} $\beta_1 = \angle \text{H}_2\text{N}_{10}\text{C}_2\text{N}_1$ and $\beta_2 = \angle \text{H}_1\text{N}_{10}\text{C}_2\text{N}_3$, for C_{1,2,4} $\beta_1 = \angle \text{H}_1\text{N}_7\text{C}_4\text{C}_5$ and $\beta_2 = \angle \text{H}_2\text{N}_7\text{C}_4\text{N}_3$. The indices referring to atoms in nucleobase structures are explained in Fig. S6 in the Supporting Information⁶⁹.

	β_1			β_2		
	PBE	KIPZ	MP2 ³⁷	PBE	KIPZ	MP2 ³⁷
A	0.05	23.16	18.70	0.05	23.15	21.10
G ₁	2.00	10.80	-	2.03	46.12	-
G ₂	2.05	17.60	11.80	2.03	35.36	43.20
G ₃	1.82	23.69	-	1.92	24.44	-
G ₄	1.81	25.35	-	1.88	22.51	-
G ₅	1.82	21.97	-	1.88	27.69	-
C ₁	8.25	30.64	26.20	7.20	17.27	14.10
C ₂	8.68	27.97	-	8.01	21.01	-
C ₄	8.90	28.57	-	8.02	20.53	-

A remarkable success of the structural optimizations done with the KIPZ functional is the correct description of the slight tilt of the amino groups of nucleobases with respect to their aromatic rings, which can be seen in A, G₁, G₂, G₃, G₄, G₅, C₁, C₂ and C₄. A correct prediction of this nonplanarity is an important step towards accurate predictions of the structure of DNA and towards the understanding of molecular recognition processes in biological systems. The presence of the amino-group tilt was subject to some controversy in the past (for a detailed discussion, see, e.g., Refs. 72 and 73). Theoretical studies of DNA and RNA nucleobases^{74,75} using the HF method with the 3-21G basis set originally suggested

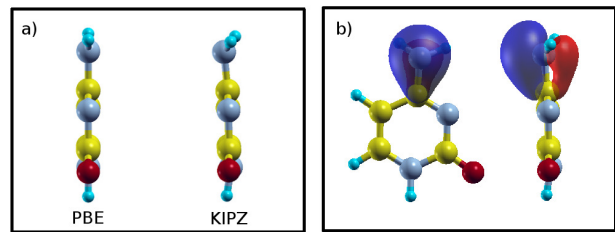


Figure 8. (a) Side view of the C₁ structures optimized using PBE and KIPZ functionals. (b) Side view and top view of C₁, with an iso-surface plot of one of the variational orbitals supporting the bond of the amino group.

Table V. IPs and EAs of dipole-bound states computed using the KIPZ functional on top of PBE (@PBE) structure or on top of the structure resulting from scf-KIPZ optimization (@KIPZ, see Fig. 1).

geo.	IP			EA		
	@PBE	@KIPZ	Expt. ^a	@PBE	@KIPZ	Expt. ^b
A	8.25	8.41	8.30–8.50	-0.01	-0.02	0.01
G ₁	7.86	8.25	8.00–8.30	0.08	0.03	–
G ₂	7.77	8.07	–	0.13	0.08	–
G ₃	7.76	8.01	–	0.02	-0.02	–
G ₄	7.79	8.02	–	0.01	0.01	–
G ₅	7.78	8.05	–	0.04	0.02	–
C ₁	8.48	8.70	8.80–9.00	0.11	0.11	0.23
C ₂	8.56	8.67	–	0.05	0.03	0.09
C ₄	8.57	8.67	–	0.03	0.05	–

^a Collected from Reference 9.

^b References 63 and 16.

the molecules to be perfectly planar. Subsequent calculations carried out at the HF level with polarized basis sets of atomic orbitals were instead able to observe a weak nonplanarity of the amino groups of the base molecules⁷⁶. Later, post-HF approaches indicated an even stronger amino-group pyramidalization. For C₁, for example, Bludsky et al⁷⁷ obtained amino group hydrogen dihedral angles of 5.5° and 21.4° using the HF/6-31G** and MP2/6-31G* levels of theory, respectively. This MP2 result is close to the predictions by Sponer and Habza of a dihedral angle of 26.2°³⁷. Our DFT-PBE calculations, in agreement with those by Di Felice et al⁷⁸, do not find significant deviations from planarity (see Table IV), with a dihedral angle of 8.25° for C₁, and even smaller angles for the other DNA bases. The strong amino-group pyramidalization can be instead reproduced very well within KIPZ, with amino hydrogen dihedral angles of the same order of the MP2 results. On panel (a) of Fig. 8, we show as an example the equilibrium geometry of the C₁ molecule predicted by KIPZ, as compared to the PBE structure. On panel (b) we show the KIPZ tilted structure with the variational orbital building the distorted π -bond that supports the connection between the amino group and the molecule. The orbital asymmetry shown is correlated to a switch from a *sp*²-like hybridization of

the N atom to a more sp^3 -like bond configuration in the tilted structure.

We find that predicting correctly the amino group shape of the nucleobases is important in view of computing accurately the electronic excitation spectrum of the molecules. In Table V we compare IPs and EAs computed with the KIPZ functional top of the PBE structure (a mostly planar molecule) and on top of the KIPZ (nonplanar) structure. Describing well the pyramidalization of NH_2 - group increases ionization energies by about 0.2–0.3 eV (with a much less pronounced effect on LUMO states), resulting in theoretical predictions that are closer to experiments. We stress that all the results of Sec. III A, III B, and III C used this scf-KIPZ approach for determining geometry.

IV. CONCLUSION

In this paper we have explored the capability of the KIPZ functional to predict both spectral properties, such as ionization potential, electron affinities, ultraviolet photoemission spectral, and geometries of DNA and RNA nucleobases and their tautomer variants, showing an excellent agreement with experiments, with mean absolute errors for the first IPs and EAs that are smaller than 0.1 eV. The accuracy of the KIPZ functional in predicting IPs and EAs of nucleobases is comparable to that of more computationally intensive methods derived from many-body perturbation theory, such as G_0W_0 and scf-GW, or quantum chemistry, such as CASPT2 and CCSD(T). In addition, for EAs the empty excited states of the nucleobases can be found in two variants: the delocalized and weakly bound DB states and the localized and unbound VB states, which are close in energy, and which make the study of these systems with localized basis set extremely

challenging, but yield very accurate results in the case of KIPZ.

Similarly, the photoemission spectra of nucleobase molecules show an excellent agreement with UPS data measured at the same incoming photon energy. These results support the suggestion (see Refs. 22 and 36) that KC functionals can be seen as a beyond-DFT approach where the spectral potential²¹, rather than the exchange-correlation one, is directly approximated, and provide both a conceptual and a practical framework to predict spectral properties from functional theories, rather than perturbative approaches. In addition, the excellent agreement between theoretical and experimental spectra allows us to assign the orbitals of the low binding energy UPS excitations using the KIPZ eigenstates, as well as to resolve the experimental spectra of C and G molecules by attributing them to a weighted spectrum of their tautomers.

By exploring self-consistent screening (scf-KIPZ) to optimize molecular geometries we find that correctly predicting the structural properties of the bases, especially for the amino groups, yields a better agreement between theoretical and experimental IP energies. Overall, we believe that our results are a step towards further studies of the electronic structure of complex DNA and RNA sequences, for which methods from many-body perturbation theory or quantum chemistry would be computationally very challenging.

ACKNOWLEDGMENTS

We acknowledge partial support from the Swiss National Centre for Computational Design and Discovery of Novel Materials (MARVEL), and the EU Centre of Excellence "MaX - Material design at the eXascale" (Grant No. 676598).

* linh.nguyen@epfl.ch

- ¹ A.-O. Colson and M. D. Sevilla, *The Journal of Physical Chemistry* **99**, 3867 (1995).
- ² N. S. Hush and A. S. Cheung, *Chemical Physics Letters* **34**, 11 (1975).
- ³ D. Dougherty, E. S. Younathan, R. Voll, S. Abdunur, and S. P. McGlynn, *Journal of Electron Spectroscopy and Related Phenomena* **13**, 379 (1978).
- ⁴ K.-W. Choi, J.-H. Lee, and S. K. Kim, *Journal of the American Chemical Society* **127**, 15674 (2005).
- ⁵ A. B. Trofimov, J. Schirmer, V. B. Kobychyev, A. W. Potts, D. M. P. Holland, and L. Karlsson, *Journal of Physics B: Atomic, Molecular and Optical Physics* **39**, 305 (2006).
- ⁶ M. Schwell, H.-W. Jochims, H. Baumgärtel, and S. Leach, *Chemical Physics* **353**, 145 (2008).
- ⁷ I. L. Zaytseva, A. B. Trofimov, J. Schirmer, O. Plekan, V. Feyer, R. Richter, M. Coreno, and K. C. Prince, *The Journal of Physical Chemistry A* **113**, 15142 (2009).
- ⁸ O. Kostko, K. Bravaya, A. Krylov, and M. Ahmed, *Phys-*

- ical Chemistry Chemical Physics* **12**, 2860 (2010).
- ⁹ D. Roca-Sanjuán, M. Rubio, M. Merchán, and L. Serrano-Andrés, *The Journal of Chemical Physics* **125**, 084302 (2006).
- ¹⁰ D. Roca-Sanjuán, M. Merchán, L. Serrano-Andrés, and M. Rubio, *The Journal of Chemical Physics* **129**, 095104 (2008).
- ¹¹ X. Qian, P. Umari, and N. Marzari, *Physical Review B* **84**, 075103 (2011).
- ¹² M. E. Foster and B. M. Wong, *Journal of Chemical Theory and Computation* **8**, 2682 (2012).
- ¹³ H. Sun, S. Zhang, C. Zhong, and Z. Sun, *Journal of Computational Chemistry* **37**, 684 (2016).
- ¹⁴ H. Satzger, D. Townsend, and A. Stolow, *Chemical Physics Letters* **430**, 144 (2006).
- ¹⁵ V. Feyer, O. Plekan, R. Richter, M. Coreno, G. Vallloera, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, T. E. Moskovskaya, E. V. Gromov, and J. Schirmer, *The Journal of Physical Chemistry A* **113**, 5736 (2009).

- ¹⁶ J. Schiedt, R. Weinkauff, D. M. Neumark, and E. W. Schlag, *Chemical Physics* **239**, 511 (1998).
- ¹⁷ K. B. Bravaya, O. Kostko, S. Dolgikh, A. Landau, M. Ahmed, and A. I. Krylov, *The Journal of Physical Chemistry A* **114**, 12305 (2010).
- ¹⁸ G. Onida, L. Reining, and A. Rubio, *Review Modern Physics* **74**, 601 (2002).
- ¹⁹ J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, *Phys. Rev. Lett.* **49**, 1691 (1982).
- ²⁰ J. P. Perdew and M. Levy, *Phys. Rev. B* **56**, 16021 (1997).
- ²¹ M. Gatti, V. Olevano, L. Reining, and I. V. Tokatly, *Phys. Rev. Lett.* **99**, 057401 (2007).
- ²² A. Ferretti, I. Dabo, M. Cococcioni, and N. Marzari, *Physical Review B* **89**, 195134 (2014).
- ²³ L. J. Sham and W. Kohn, *Phys. Rev.* **145**, 561 (1966).
- ²⁴ M. E. Casida, *Phys. Rev. A* **51**, 2005 (1995).
- ²⁵ D. Chong, O. Gritsenko, and E. Baerends, *J. Chem. Phys.* **116**, 1760 (2002).
- ²⁶ I. Dabo, *Towards First-Principles Electrochemistry*, PhD thesis, MIT, Cambridge MA, 2008, <http://dspace.mit.edu/handle/1721.1/44320>.
- ²⁷ I. Dabo, M. Cococcioni, and N. Marzari, *arXiv:0901.2637v1* (2009).
- ²⁸ I. Dabo, A. Ferretti, N. Poilvert, Y. Li, N. Marzari, and M. Cococcioni, *Physical Review B* **82**, 115121 (2010).
- ²⁹ I. Dabo, A. Ferretti, G. Borghi, N. L. Nguyen, N. Poilvert, C. H. Park, M. Cococcioni, and N. Marzari, *Psi-K Newsletter* **119**, 1 (2013).
- ³⁰ I. Dabo, A. Ferretti, and N. Marzari, *Topics in Current Chemistry* **347**, 193 (2014).
- ³¹ I. Dabo, A. Ferretti, C.-H. Park, N. Poilvert, Y. Li, M. Cococcioni, and N. Marzari, *Physical Chemistry Chemical Physics* **15**, 685 (2013).
- ³² G. Borghi, A. Ferretti, N. L. Nguyen, I. Dabo, and N. Marzari, *Physical Review B* **90**, 075135 (2014).
- ³³ H. J. Kulik, M. Cococcioni, D. A. Scherlis, and N. Marzari, *Physical Review Letters* **97**, 103001 (2006).
- ³⁴ H. J. Kulik and N. Marzari, *The Journal of Chemical Physics* **129**, 134314 (2008).
- ³⁵ M. Cococcioni and S. de Gironcoli, *Physical Review B* **71**, 035105 (2005).
- ³⁶ N. L. Nguyen, G. Borghi, A. Ferretti, I. Dabo, and N. Marzari, *Physical Review Letters* **114**, 166405 (2015).
- ³⁷ J. Šponer and P. Hobza, *The Journal of Physical Chemistry* **98**, 3161 (1994).
- ³⁸ C. Filippi, C. J. Umrigar, and M. Taut, *The Journal of Chemical Physics* **100**, 1290 (1994).
- ³⁹ J. P. Perdew and A. Zunger, *Physical Review B* **23**, 5048 (1981).
- ⁴⁰ J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
- ⁴¹ A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Science* **321**, 792 (2008).
- ⁴² O. A. Vydrov, G. E. Scuseria, J. P. Perdew, A. Ruzsinszky, and G. I. Csonka, *The Journal of Chemical Physics* **124**, 94108 (2006).
- ⁴³ S. Klüpfel, P. Klüpfel, and H. Jónsson, *The Journal of Chemical Physics* **137**, 124102 (2012).
- ⁴⁴ E. Kraissler and L. Kronik, *Physical Review Letters* **110**, 126403 (2013).
- ⁴⁵ O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, *The Journal of Chemical Physics* **126**, 154109 (2007).
- ⁴⁶ J. C. Slater, *Quantum Theory of Molecules and Solids: The Self-Consistent Field for Molecules and Solids*, volume 4 of *International Series in Pure and Applied Physics*, McGraw-Hill, New York, 1974.
- ⁴⁷ D. Hofmann, S. Klüpfel, P. Klüpfel, and S. Kümmel, *Physical Review A* **85**, 062514 (2012).
- ⁴⁸ S. Lehtola and H. Jónsson, *Journal of Chemical Theory and Computation* **10**, 5324 (2014).
- ⁴⁹ G. Borghi, C.-H. Park, N. L. Nguyen, A. Ferretti, and N. Marzari, *Phys. Rev. B* **91**, 155112 (2015).
- ⁵⁰ N. Marzari, D. Vanderbilt, and M. C. Payne, *Physical Review Letters* **79**, 1337 (1997).
- ⁵¹ M. R. Pederson, R. A. Heaton, and C. C. Lin, *The Journal of Chemical Physics* **80**, 1972 (1984).
- ⁵² T. Stein, H. Eisenberg, L. Kronik, and R. Baer, *Phys. Rev. Lett.* **105**, 266802 (2010).
- ⁵³ P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, *J Phys-Condens Mat* **21**, 395502 (2009).
- ⁵⁴ Y. Li and I. Dabo, *Physical Review B* **84**, 155127 (2011).
- ⁵⁵ The norm-conserving pseudopotentials for C, O, N and H were taken from Quantum ESPRESSO pseudopotential download page: <http://www.quantum-espresso.org/pseudo.php>.
- ⁵⁶ S. Klüpfel, P. Klüpfel, and H. Jónsson, *Physical Review A* **84**, 050501 (2011).
- ⁵⁷ S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. Orozco, F. J. Luque, J. Sponer, P. Slavicek, and P. Hobza, *Phys. Chem. Chem. Phys.* **4**, 4192 (2002).
- ⁵⁸ O. Plekan, V. Feyer, R. Richter, M. Coreno, G. Valllosera, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, T. E. Moskovskaya, E. V. Gromov, and J. Schirmer, *The Journal of Physical Chemistry A* **113**, 9376 (2009).
- ⁵⁹ C. Faber, C. Attaccalite, V. Olevano, E. Runge, and X. Blase, *Physical Review B* **83**, 115123 (2011).
- ⁶⁰ M. Haranczyk and M. Gutowski, *Journal of the American Chemical Society* **127**, 699 (2005).
- ⁶¹ D. Dougherty, E. Younathan, R. Voll, S. Abdunur, and S. McGlynn, *Journal of Electron Spectroscopy and Related Phenomena* **13**, 379 (1978).
- ⁶² J. H. Hendricks, S. A. Lyapustina, H. L. d. Clercq, J. T. Snodgrass, and K. H. Bowen, *The Journal of Chemical Physics* **104**, 7788 (1996).
- ⁶³ C. Desfrancois, H. Abdoul-Carime, and J. P. Schermann, *The Journal of Chemical Physics* **104**, 7792 (1996).
- ⁶⁴ T. J. O'Donnell, P. R. LeBreton, J. D. Petke, and L. L. Shipman, *The Journal of Physical Chemistry* **84**, 1975 (1980).
- ⁶⁵ K. Aflatooni, G. A. Gallup, and P. D. Burrow, *The Journal of Physical Chemistry A* **102**, 6205 (1998).
- ⁶⁶ A. Damascelli, Z. Hussain, and Z.-X. Shen, *Rev. Mod. Phys.* **75**, 473 (2003).
- ⁶⁷ J. W. Rabalais, T. P. Debies, J. L. Berkosky, J. J. Huang, and F. O. Ellison, *The Journal of Chemical Physics* **61**, 516 (1974).
- ⁶⁸ M. Walter and H. Häkkinen, *New Journal of Physics* **10**, 043018 (2008).
- ⁶⁹ The Supporting Information contains the UPS spectra of the molecules as function of the incoming photon energies (Figs. S1 and S2), the iso-density plots and binding

- energies of the orbitals close to HOMO (Figs. S3, S4 and S5), bond length and angle values for A, T and U presented in Table III (Table S1, S2 and S3) and the index of the atoms of the dihedral angles presented in Table IV (Fig. S6).
- ⁷⁰ S. Refaely-Abramson, S. Sharifzadeh, N. Govind, J. Autschbach, J. B. Neaton, R. Baer, and L. Kronik, *Physical Review Letters* **109**, 226405 (2012).
- ⁷¹ L. Clowney, S. C. Jain, A. R. Srinivasan, J. Westbrook, W. K. Olson, and H. M. Berman, *Journal of the American Chemical Society* **118**, 509 (1996).
- ⁷² P. Hobza and J. Šponer, *Chemical Reviews* **99**, 3247 (1999).
- ⁷³ J. Šponer and P. Hobza, *International Journal of Quantum Chemistry* **57**, 959 (1996).
- ⁷⁴ N. V. Riggs, *Chemical Physics Letters* **177**, 447 (1991).
- ⁷⁵ R. D. Brown, P. D. Godfrey, D. McNaughton, and A. P. Pierlot, *Chemical Physics Letters* **156**, 61 (1989).
- ⁷⁶ J. Leszczyński, *International Journal of Quantum Chemistry* **44**, 43 (1992).
- ⁷⁷ O. Bludský, J. Šponer, J. Leszczynski, V. Špirko, and P. Hobza, *The Journal of Chemical Physics* **105**, 11042 (1996).
- ⁷⁸ R. Di Felice, A. Calzolari, E. Molinari, and A. Garbesi, *Physical Review B* **65**, 045104 (2001).